


IT'S RAINING CREDs: CRAWLING DOCKERHUB FOR LEAKED SECRETS AT SCALE

Aliz Hammond, 2023
BSides Singapore, 22nd September 2023



About me

- Aliz Hammond
- Background in bug hunting
 - Mostly binary-level
- Detection of weird malware techniques
 - Often at ring0
- Recently (a year or so ago) moved towards web stuff
- Working at watchTowr
 - Alongside red-teamers with very practical experience



We are an **Attack Surface Management technology** company, **built by red teamers**.

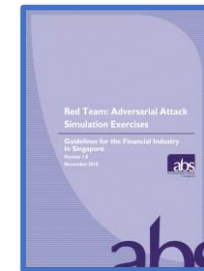
As a team, we have **years of experience** breaking into enterprises, **simulating ransomware gangs and the North Koreans**.



HONG KONG MONETARY AUTHORITY
香港金融管理局



DeNederlandscheBank





Cutting-edge offensive security research



FORTINET®

 **OPENVPN**®



 **ForgeRock**®

orbeon 

 **Confluence**

SONICWALL®

JUNIPER®
NETWORKS

We see recurring patterns

Neglected and unintentionally extended attack surfaces

BREACH EXPLAINED

GitHub Exposed A Private SSH Key: What You Need To Know

Everyone has secrets leakage incidents from time to time, even massive players like GitHub. This is a good reminder we all need to stay vigilant and embrace the right tools to help us stay safe.

students' grades and personal info

Educator gets an F for security

Jessica Lynn Hardcastle

Twilio: Someone waltzed into our unsecured AWS S3 silo, added dodgy code to our JavaScript SDK for customers

API dev kit remained modified for hours, says source

Microsoft leaks 38TB of private data via unsecured Azure storage

By Sergiu Gatlan

September 18, 2023 11:18 AM 2

exposed 100,000

accidentally leaked dozens of terabytes of sensitive data starting in source AI learning models to a public GitHub repository.

discovered by cloud security firm Wiz whose security researchers found and publicly shared the URL for a misconfigured Azure Blob storage bucket

S3 bucket mess up exposed 182GB of senior US, Canada citizens data


The misconfigured S3 bucket was owned by SeniorAdvisor, a consumer ratings, and reviews website.

BY WAQAS · AUGUST 13, 2021 · 2 MINUTE READ

Perhaps **DockerHub** contains
something interesting?



It's raining



[Explore](#)
[Pricing](#)
[Sign In](#)
[Register](#)

Filters

1 - 25 of 10,000 available results.

Suggested


Products


☐ Images


☐ Extensions

☐ Plugins

Trusted Content

☐  Docker Official Image ⓘ

☐  Verified Publisher ⓘ

☐  Sponsored OSS ⓘ

Operating Systems

☐ Linux

☐ Windows

Architectures

☐ ARM

☐ ARM 64


☐ IBM POWER


☐ IBM Z

☐ PowerPC 64 LE

☐ x86

☐ x86-64



alpine  DOCKER OFFICIAL IMAGE · 1B+ · 9.8K


Updated 14 days ago

A minimal Docker image based on Alpine Linux with a complete package index and onl...


Linux IBM Z riscv64 x86-64 ARM ARM 64 386 PowerPC 64 LE


Pulls: 8,577,848

Last week



[Learn more](#)



nginx  DOCKER OFFICIAL IMAGE · 1B+ · 10K+


Updated 3 days ago

Official build of Nginx.


Linux IBM Z x86-64 ARM ARM 64 386 mips64le PowerPC 64 LE


Pulls: 31,751,671

Last week



[Learn more](#)



busybox  DOCKER OFFICIAL IMAGE · 1B+ · 2.9K


Updated 11 days ago

Busybox base image.


Linux ARM ARM 64 386 mips64le PowerPC 64 LE riscv64 IBM Z x86-64


Pulls: 12,408,315

Last week



[Learn more](#)



ubuntu  DOCKER OFFICIAL IMAGE · 1B+ · 10K+


Updated 12 days ago

Ubuntu is a Debian-based Linux operating system based on free software.

Linux IBM Z 386 riscv64 x86-64 ARM ARM 64 PowerPC 64 LE

Pulls: 22,145,666

Last week



[Learn more](#)

It's raining creds

A bit of quick Python, and....


```
{
  "cf_email": "<watchtower redacted>","",
  "cf_apikey": "<watchtower redacted>",
  "cf_org": "<watchtower redacted>Prod",
  "cf_space": "<watchtower redacted>prodspace2",
  "cf_broker_memory": "512M",

  "devex": {
    "ace_app_space": "opsconsole",
    "ace_app_suffix": "dev",
    "bssr_client_id": "<watchtower redacted>",
    "bssr_client_secret": "<watchtower redacted>",
    "node_env": "production",
    "session_key": "opsConsole.sid",
    "session_secret": "<watchtower redacted>",
    "slack_endpoint": "<TBD>",
    "uaa_callback_url": "<watchtower redacted>",
    "uaa_client_id": "<watchtower redacted>",
    "uaa_client_secret": "<watchtower redacted>"
  },
  "CLOUDANT_USER": "<watchtower redacted>",
  "CLOUDANT_PASSWORD": "<watchtower redacted>",
  "BLUEMIX_CLOUDANT_DB_NAME": "<watchtower redacted>",
  "TERRAFORM_CLOUDANT_DB_NAME": "<watchtower redacted>",
  "CLOUDANT_DB_URL": "<watchtower redacted>",
  "BLUEMIX_SPACE": "<watchtower redacted>",
  "UAA_CLIENT_SECRET": "<watchtower redacted>",
  "UAA_CALLBACK": ""<watchtower redacted>","",
  "SESSION_CACHE": "Redis-cam-ui",
  "TERRAFORM_PLUGIN_ENDPOINTURL": "<watchtower redacted>/api",
  "CAM_TOKEN": "<watchtower redacted>",
  "A8_REGISTRY_TOKEN": "<watchtower redacted>",
  "A8_CONTROLLER_TOKEN": "<watchtower redacted>",
  "REDIS_HOST": "<watchtower redacted>",
  "REDIS_PASSWORD": "<watchtower redacted>",
  "REDIS_PORT": "<watchtower redacted>",
  "SRE_SLACK_WEBHOOK_URI": "https://hooks.slack.com/services/<watchtower redacted>",
}
```

Time to do this **At Scale**TM

It's raining creds

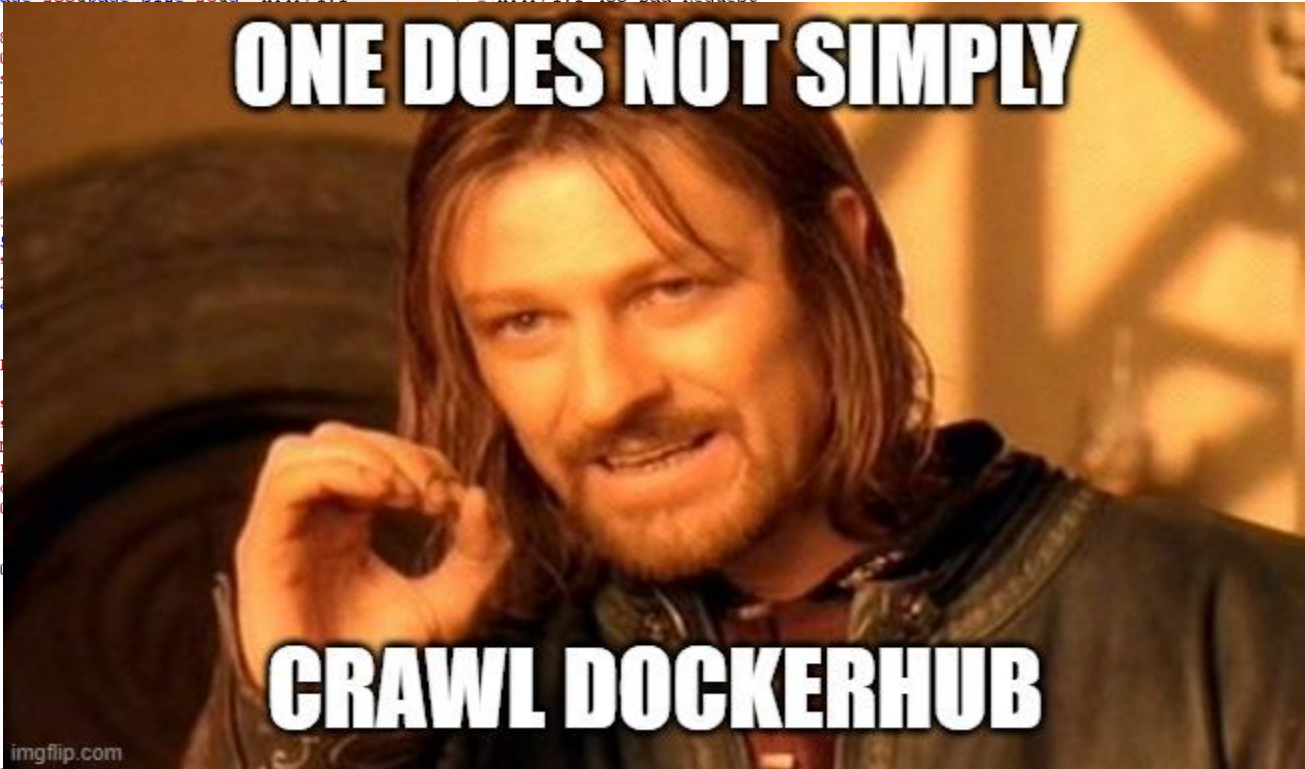
Overall system design

- Crawl hub.docker.com
- Extract, dedup files from Docker containers
- Store file metadata in MySQL
- Run some kind of analysis phase to locate secrets

First step: **Crawl** DockerHub

ⓘ ⚙ ⬅ ➡

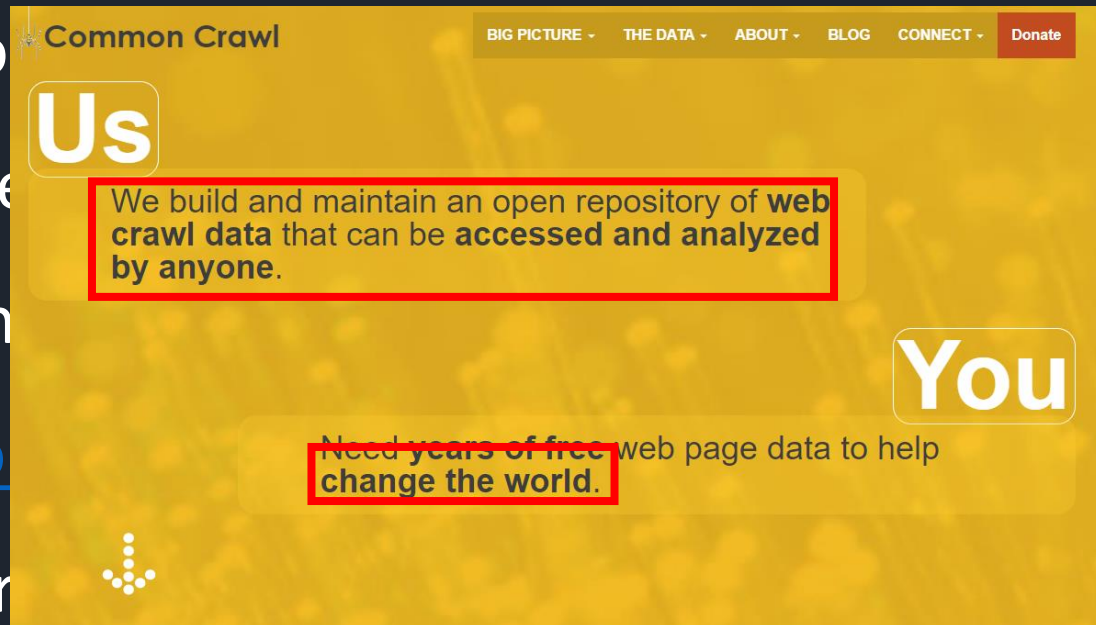
Cr



It's raining creds

Crawling

- One way – dictionary
- Slow, inefficient
- What we really need
- Enter: <https://commoncrawl.org/>
- Public, free crawl



[commoncrawl.com](https://commoncrawl.org/)

It's raining creds

Crawling

```
select count(*)  
FROM "ccindex"."ccindex"  
WHERE crawl like 'CC-MAIN-2022-33'  
and url_host_name = 'hub.docker.com'
```

Over 20,000 Docker images

Second step: Download **thousands** of
containers

It's raining creds

Fetching

- Could just 'docker pull'
 - Need to flush periodically to free disk space
 - Need to examine every file in every container

```
FROM ubuntu:jammy  
RUN adduser foo
```

```
FROM ubuntu:jammy  
RUN adduser bar
```

It's raining creds

Fetching

- Need to be layer-aware
 - Index Ubuntu once
 - Index changes from first container
 - Index changes from second container

```
FROM ubuntu:jammy  
RUN adduser foo
```

```
FROM ubuntu:jammy  
RUN adduser bar
```

It's raining creds

Fetching

- Fetch container manifest via HTTP API
 - This contains list of layers
- Fetch each layer we haven't seen before
 - They're just .tgz so decompress in memory

It's raining creds

Fetching

- DockerHub aggressively rate limits
 - Disposable web proxies

It's raining creds

Fetching

- Becomes fairly standard at this point
 - One database node
 - Number of 'spider' nodes (5)
 - Each with a number of HTTP proxies (5)
 - Fileserver to store files themselves

Third step: Storing **millions** of files

It's raining creds

Finding secrets

“Artifacts”

artifacts	
id	INT
filename	VARCHAR(255)
pathid	INT
mode	MEDIUMINT
hashID	INT
filesize	INT
ownerUID	INT
groupUID	INT
perm_world_r	TINYINT(1)
perm_world_w	TINYINT(1)
perm_world_x	TINYINT(1)
perm_group_r	TINYINT(1)
perm_group_w	TINYINT(1)
perm_group_x	TINYINT(1)
perm_owner_r	TINYINT(1)
perm_owner_w	TINYINT(1)
perm_owner_x	TINYINT(1)
perm_sticky	TINYINT(1)
perm_guid	TINYINT(1)
perm_suid	TINYINT(1)
Indexes	

```
create table artifacts(
  `mode` mediumint UNSIGNED NOT NULL,
  `perm_world_r` bool AS (mode&0x001!=0),
  `perm_world_w` bool AS (mode&0x002!=0),
  `perm_world_x` bool AS (mode&0x004!=0),
  `perm_group_r` bool AS (mode&0x008!=0),
  `perm_group_w` bool AS (mode&0x010!=0),
  `perm_group_x` bool AS (mode&0x020!=0),
  `perm_owner_r` bool AS (mode&0x040!=0),
  `perm_owner_w` bool AS (mode&0x080!=0),
  `perm_owner_x` bool AS (mode&0x100!=0),
  `perm_sticky` bool AS (mode&0x200!=0),
  `perm_guid` bool AS (mode&0x400!=0),
  `perm_suid` bool AS (mode&0x800!=0),
```


Results: 134 million files over 22,217
containers

It's raining creds

Finding secrets

```
mysql> select filename, count(distinct(hash)) from dockerArtifacts where filename in ( 'id_rsa', 'id_dsa', 'id_ed25519', 'id_ecdsa' ) group by filename;
```

filename	count(distinct(hash))
id_dsa	15
id_ecdsa	5
id_ed25519	16
id_rsa	117

4 rows in set (1.90 sec)

```
mysql> select path, filename, hash from artifacts join artifactHashes on artifactHashes.id = artifacts.hashID join artifact_path on artifact_path.id = artifacts.pathid where path like '%.ethereum' limit 10;
```

path	filename	hash
/root/.ethereum	nodekey	b8ebd8a276902106ebe6427b11353d3d876f1203
/root/.ethereum	genesis.json	58eb70c3fe0d24d1ce044c34d19f6da49d69699d
/root/.ethereum	passwords	0f4ae616a5e9520fdcb63b1103813861eed28f34
/root/.ethereum	static-nodes.json	82fe1224835ba83dead3d75997b5862869c571e4
/root/.ethereum	genesis.json	8fad30095f55b5824fd4b0e1e7ab6d5992d4c78b
/root/.ethereum	bank1.id	71681d513935fec9e0c0fce5ffbd4153df54f388
/root/.ethereum	bank2.id	b7be0d26069769fc57ce7d076f6b7c24c589a13e
/root/.ethereum	buyer.id	558515bc23a778c40e2ce76e11f1b7afbb92586f
/root/.ethereum	carrier.id	71de027c765d3fa0d85a75bb44427e4b8a18b5e8
/root/.ethereum	genesis.json	67ce05b2451cd6e59d279131d5e7cf893d89ea90

10 rows in set (13.88 sec)

It's raining

Finding secrets

```
mysql> select count(distinct(hashID)) from artifacts where filename = '.env';
+-----+
| count(distinct(hashID)) |
+-----+
| 194 |
+-----+
1 row in set (0.00 sec)
```

```
root@ubuntu:/var/lib# cat /docker/artifacts/ad4bb1ca8e4aeeb69c449b813d6e19ec78cd45e
APP_ENV=production
APP_DEBUG=true
APP_KEY=base64:m[REDACTED]ahk=
APP_URL=http://api.[REDACTED].com
DB_CONNECTION=pgsql
DB_HOST=db
DB_PORT=5432
DB_DATABASE=bizmerk
DB_USERNAME=root
DB_PASSWORD=t[REDACTED]*1W
CACHE_DRIVER=file
SESSION_DRIVER=file
QUEUE_DRIVER=sync
REDIS_HOST=127.0.0.1
REDIS_PASSWORD=null
REDIS_PORT=6379
MAIL_DRIVER=mailgun
MAIL_HOST=smtp.mailgun.org
MAIL_PORT=25
MAIL_USERNAME=postmaster@[REDACTED].com
MAIL_PASSWORD=8b89[REDACTED]72770
MAIL_ENCRYPTION=tls
MAILGUN_DOMAIN=[REDACTED].com
MAILGUN_SECRET=key-00672a6[REDACTED]a6
FRONT_URL=http://app.[REDACTED].com
```

It's raining creds

Finding secrets

- Some analysis possible already
 - Filename
 - File permissions
- Context missing from artifacts
 - Usually impossible to verify secrets
 - Exception: host keys

It's raining creds

Finding secrets

- Host keys can be found via Shodan
 - 724 unique IP addresses

It's raining creds

Finding secrets

```
mysql> select concat(path, '/', filename), count(*) from artifactsOnline join artifacts on artifactsOnline.fileid = artifacts.id join artifact_path on artifact_path.id = artifacts.pathid group by concat(path, '/', filename) order by concat(path, '/', filename) ;
```

concat(path, '/', filename)	count(*)
/app/vendor/phpseclib/phpseclib/tests/Unit/File/X509/CSRTTest.php	14
/etc/ssh/ssh_host_rsa_key	58
/home/ap-loadtester/apenv/site-packages/twisted/conch/manhole_ssh.pyc	3
/pentest/exploitation/badkeys/host/Actiontec_q2000_rsa.key	100
/pentest/exploitation/badkeys/host/Seagate_GoFlex_rsa.key	100
/pentest/exploitation/badkeys/host/Tplink_tdw8960n-V1_rsa.key	14
/pentest/exploitation/badkeys/host/Tplink_w8950nd_rsa.key	74
/pentest/exploitation/badkeys/host/Zhone_6512a1_rsa.key	32
/pentest/exploitation/badkeys/host/Zyxel_p870h_rsa.key	10
/pentest/exploitation/badkeys/host/Zyxel_pmg1006_rsa.key	100
/pentest/exploitation/badkeys/host/zyxel-vmg1312_rsa.key	100
/usr/lib/python2.7/site-packages/synapse/util/manhole.py	5
/usr/lib/python2.7/site-packages/synapse/util/manhole.pyc	5
/usr/local/lib/python2.7/dist-packages/Twisted-14.0.0-py2.7-linux-x86_64.egg/twisted/conch/manhole_ssh.pyc	5
/usr/local/lib/python2.7/dist-packages/Twisted-15.5.0-py2.7-linux-x86_64.egg/twisted/conch/manhole_ssh.pyc	5
/usr/local/lib/python3.7/site-packages/synapse/util/manhole.py	2
/usr/local/reportserver/apache-tomcat-8.0.36/webapps/restserver/tmpres/hostkey.pem	82
/usr/share/clearwater/homestead/eggs/Twisted-12.3.0-py2.7-linux-x86_64.egg/twisted/conch/manhole_ssh.py	5
/usr/share/clearwater/homestead/eggs/Twisted-12.3.0-py2.7-linux-x86_64.egg/twisted/conch/manhole_ssh.pyc	5
/usr/share/doc/python-paramiko/examples/test_rsa.key	18
/usr/share/doc/python-twisted-conch/examples/sshsimpleserver.py	5

It's raining creds

Finding secrets

```
<?php
/**
 * @author    Jim Wigginton <terrafrost@php.net>
 * @copyright 2014 Jim Wigginton
 * @license   http://www.opensource.org/licenses/mit-license.html MIT License
 */
```

```
mysql> select concat(path, '/', filename) as path, group by concat(path, '/')
+-----+
| concat(path, '/', filename) |
+-----+
| /app/vendor/phpseclib/phpseclib/test |
| /etc/ssh/ssh_host_rsa_key |
| /home/ap-loadtester/apenv/site-packa |
| /usr/lib/python2.7/site-packages/syn |
| /usr/lib/python2.7/site-packages/synv |
| /usr/local/lib/python2.7/dist-packag |
| /usr/local/lib/python2.7/dist-packag |
| /usr/local/lib/python3.7/site-packag |
| /usr/local/reportserver/apache-tomca |
| /usr/share/clearwater/homestead/eggs |
| /usr/share/clearwater/homestead/eggs |
| /usr/share/doc/python-paramiko/examp |
| /usr/share/doc/python-twisted-conch/ |
+-----+

require_once 'File/X509.php';

class Unit_File_X509_CSRTest extends PhpseclibTestCase
{
    public function testLoadCSR()
    {
        $test = '-----BEGIN CERTIFICATE REQUEST-----
MIIBWzCBxQIBADAeMRwwGgYDVQQKDBNwaHBzZWNSaWgZGVtbyBjZXJ0MIGdMAsG
CSqGSIB3DQEBAQOBjQAwgYkCgYEAthDb4zoUyiRYsJ5PZrF/IJKAF9ZoHRpTxMA8
a7iyFdsl/vvZLNPsNnFTXXnGdvsyFDEsF7AubaIXw8UKFPYqQRTzSVsvnNgIoVYj
tTAXlB4oHipr7Kxcn4CXfmR0TYogyLvVZSZJYxh+CAuG4V9XM4HqkeE5gyB0sKGy
5FUU8zMCAwEAAaAAMA0GCSqGSIB3DQEBAQUAA4GBAJjdaA9K9DN5xvSi0LCmmV1E
npzHkI1Traveu0gtRjT/EzHoqjCBI0ekCZ9+fhrex8Sm6Ns9IghYyrqnE+PQko
4Nf2w2U3DWxU26D5E9DlI+bLy0Cq4jqATLjHyyAsOZY/2+U73AZ82MJM/mGdh5fQ
5RwaQHmQEzHofTzF7I+
-----END CERTIFICATE REQUEST-----';

        $x509 = new File_X509();

        $spkac = $x509->loadCSR($test);

        $this->assertInternalType('array', $spkac);
    }
}

~
~
(END)
```

It's raining creds

Finding secrets

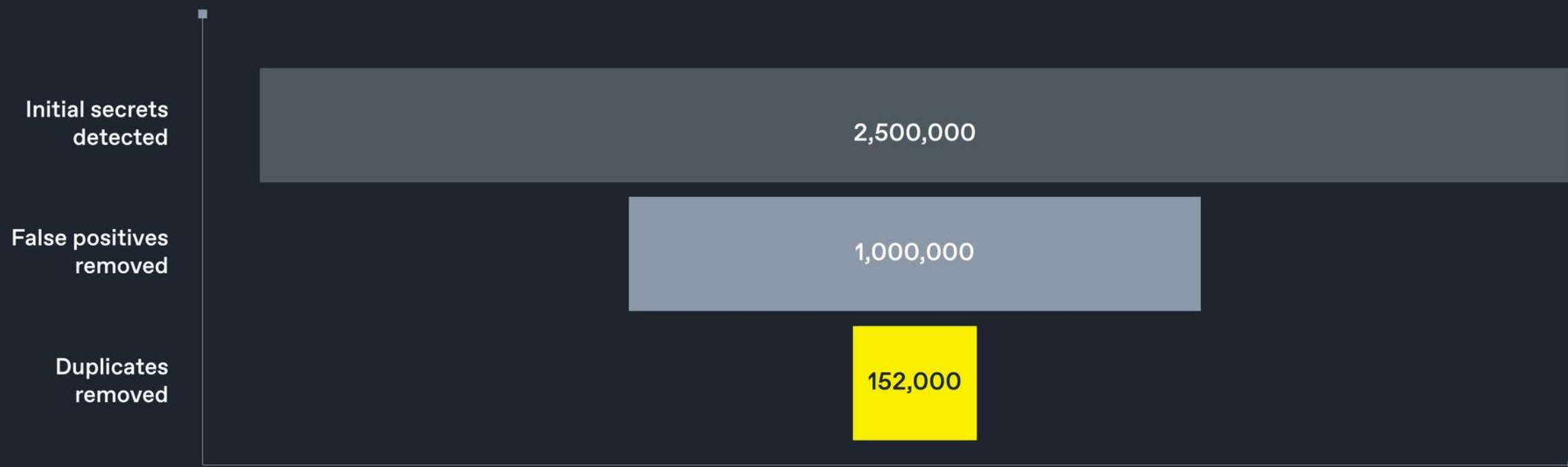
- Better analysis requires examination of file contents
- Fork opensource tool 'GitLeaks'
 - Fast, parallisable regex tool
 - Added features to deal with large dataset
- Again, stored results in MySQL

It's raining creds

Finding secrets

id	description	startLine	endLine	startColumn	endColumn	match	secret	fileid	isWellKnown
68077	Stripe	42	42	17	48	sk_live	[REDACTED]	89844802	0
68190	Stripe	137	137	22	53	sk_live	[REDACTED]	89845040	0
68251	Stripe	34	34	45	66	sk_live	[REDACTED]	89964814	0
68252	Stripe	34	34	45	66	sk_live	[REDACTED]	42762992	0
68278	Stripe	137	137	22	53	sk_live	[REDACTED]	89845039	0
68287	Stripe	9	9	137	159	sk_live	[REDACTED]	89964815	0
76	Generic API Key	25	25	15	50	auth_consumer_key="0685	[REDACTED]22"	38831817	0
77	Generic API Key	5	6	2	1	password = 03evn	[REDACTED]Pw0	19455700	0
78	Generic API Key	10	11	2	1	password = 03evn	[REDACTED]Pw0	19455700	0

So, what did we **find**?



Pruning the dataset

```
{
  "development" : {
    "accessKeyId": "AKIA[REDACTED]",
    "secretAccessKey": "[REDACTED]ACfel5cRa",
    "region": "ap-southeast-1",
    "bucket": "[REDACTED]-dev",
    "cloudfront" : "http://s3-ap-southeast-1.amazonaws.com/"
  },
}
```

```
NPM_CONFIG_LOGLEVEL=warn
PORT=5000
```

```
# CSRF Token
```

```
COOKIE_SECRET=[REDACTED]a16a4b5d6990948
7be9aab2ecdc9f[REDACTED]
```

```
# Integrated Services
```

```
MANDRILL_API_KEY=NY8[REDACTED]
CLOUDINARY_URL=cloudinary://[REDACTED]:[REDACTED]ACfel5cRa",
GOOGLE_SERVER_API_KEY= AIzaS[REDACTED]
GOOGLE_BROWSER_API_KEY=AIzaS[REDACTED]
amazonaws.com/"
```

```
# Mongolab
```

```
MONGOLAB_URI=mongodb://[REDACTED]:[REDACTED]@ds[REDACTED]
[REDACTED].com:15892/[REDACTED]
```

```
# AWS
```

```
AWS_ACCESS_KEY_ID=AKIAJT[REDACTED]
AWS_ACCESS_KEY_SECRET=vE[REDACTED]
```

```
# Keystone
```

```
[REDACTED]Uws5Lkv",
```

```
amazonaws.com/"
```

```
[REDACTED]ACfel5cRa",
```

```
amazonaws.com/"
```

```
g|
```

```
'AKIA[REDACTED]
ey = '[REDACTED]7DPiAeM'
dancy storage.
dancy = true
```

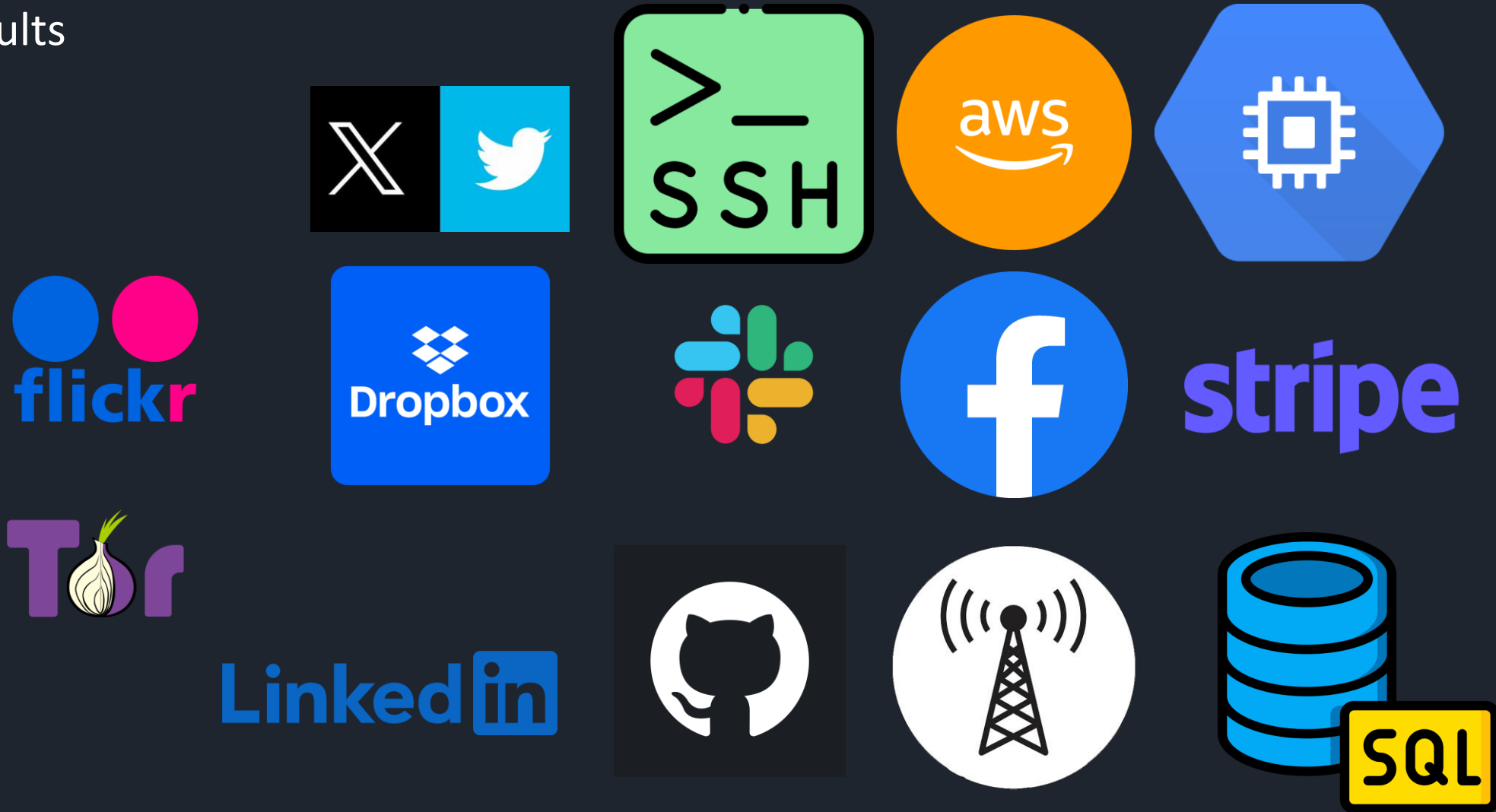
It's raining creds

Results



It's raining creds

Results



It's raining creds

While we beat the North Koreans at their own game...

```
mysql> select ownername, imagename, tagname, path, filename from dockerArtifacts where filename = 'wallet.dat';
```

ownerName	imageName	tagName	path	filename
			/root/.Vcash/data	wallet.dat
			/root/.bitcoin/gcoin	wallet.dat
			/var/lib/pivx1	wallet.dat
			/images	wallet.dat
			/images	wallet.dat

```
5 rows in set (0.05 sec)
```

Finding secrets limited only by our
imagination

It's raining creds

Summary

- Be aware of shifting attack surface
 - A tale as old as time – S3 buckets, GitHub repos, etc..
- Scale is hard
 - Scanning 10 containers is easy, scanning thousands in parallel isn't

It's raining creds

We did a whole series on this!



"I don't need no zero-days"

Docker Container Images (1/4)

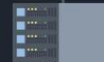
Attack Surface Research



Learning To Crawl (For DockerHub Enthusiasts, Not Toddlers)

Docker Container Images (3/4)

Attack Surface Research



All Around The World:

The Common Crawl Dataset

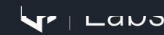
Attack Surface Research



Layer Cake: How Docker Handles Filesystem Access

Docker Container Images (2/4)

Attack Surface Research



What Does This Key Open? Detecting Secrets

Docker Container Images (4/4)

Attack Surface Research





Questions?

labs.watchtowr.com

twitter.com/watchtowrcyber

twitter.com/AlizTheHax0r